

APPLYING CORRELATION COEFFICIENTS — EDUCATIONAL ATTAINMENT AND UNEMPLOYMENT

Activity Item

The following item is part of this activity and appears at the end of this student version.

- Item 1: Unemployment Rates by Educational Attainment, 2014

Student Learning Objectives

- I will be able to make scatter plots of data.
- I will be able to calculate correlation coefficients using technology and by hand.
- I will be able to assess the strength of the linear relationship using correlation coefficients.
- I will be able to determine the impact of an influential point on the correlation coefficient.

NAME: _____ DATE: _____

Part 1 – Consider the Data

1. Examine **Item 1: Unemployment Rates by Educational Attainment, 2014.**
 - a. What is one thing you notice? Which data values catch your attention?

 - b. Write one question about the data that you could investigate.

2. Share your responses to questions 1a and 1b with your group.
 - a. Were they similar or different?

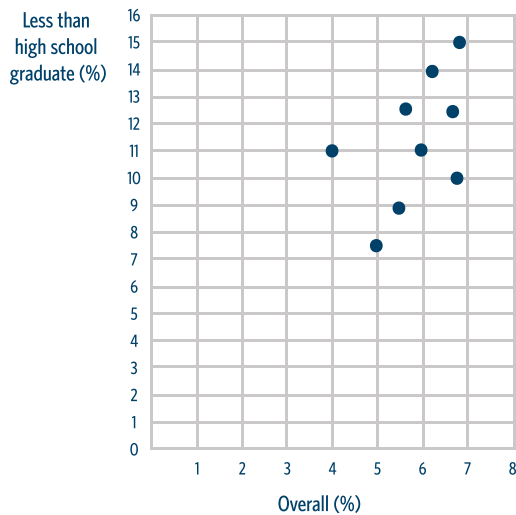
 - b. Choose two interesting questions that your group could investigate.

Part 2 – Analyze the Data

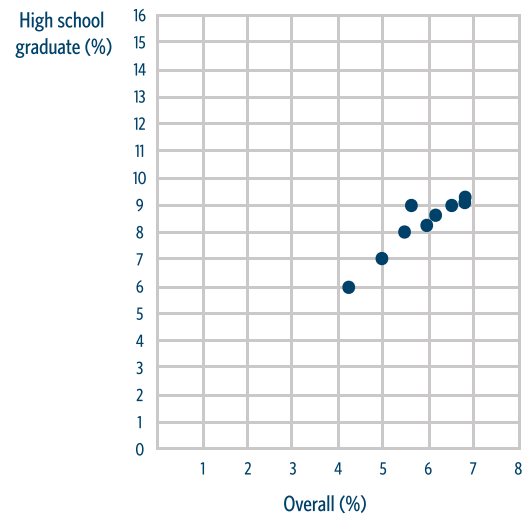
- The U.S. Census Bureau divides the United States into four regions, which break down further into nine divisions that are each represented by a dot in the scatter plots below. These graphs compare each division's unemployment rate among people of different education levels with that division's overall unemployment rate in 2014. For example, in the top left graph, the point at (4.2, 10.6) is for a Census Bureau division that had an overall unemployment rate of 4.2 percent but whose unemployment rate among those who did not graduate from high school was 10.6 percent.

Divisions' Unemployment Rates: Overall vs. Educational Attainment Levels, 2014

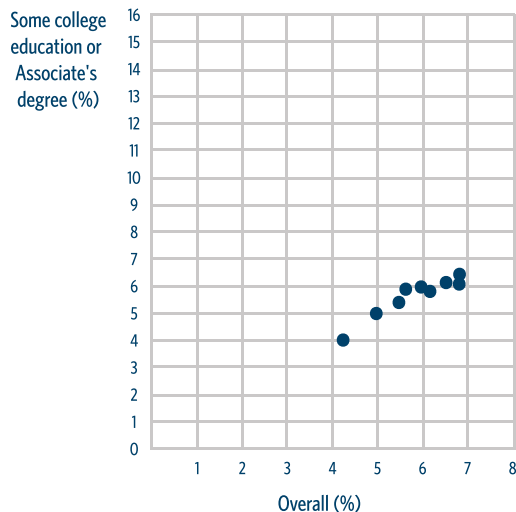
Divisions' Unemployment Rates:
Overall vs. Less Than High School Graduate



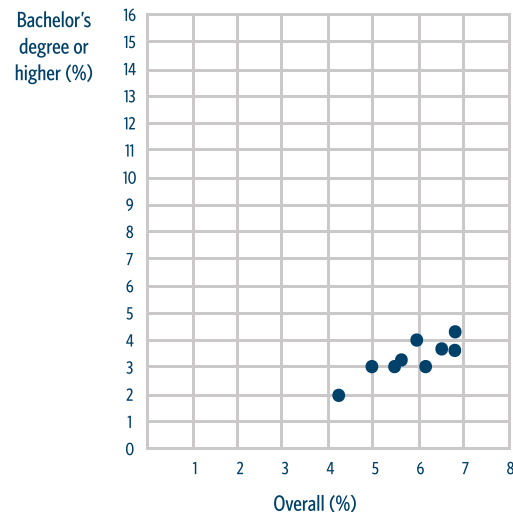
Divisions' Unemployment Rates:
Overall vs. High School Graduate



Divisions' Unemployment Rates:
Overall vs. Some College Education or Associate's Degree



Divisions' Unemployment Rates:
Overall vs. Bachelor's Degree or Higher



- a. Which two education levels' unemployment rates appear to be most strongly associated with the overall unemployment rates in a division? Explain.

- b. Which education level's unemployment rates appear to be least strongly associated with the overall unemployment rates in a division? Explain.

- c. The correlation coefficient is a measure of the strength of the linear relationship between two variables. The closer the correlation coefficient is to 0, the weaker the linear relationship. With this in mind, match each of the following correlation coefficients with the correct scatter plot from earlier. (Note: Your answers for the highest two coefficients can be interchangeable.)
 - 0.54:

 - 0.81:

 - 0.94:

 - 0.95:

2. The correlation coefficient, represented by r , is calculated using this formula:

$$r = \frac{\sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

n = the number of data pairs

x_i and y_i = specific data points

\bar{x} and \bar{y} = the means of the x and y values

$\sum_{i=1}^n$ = find the sum of everything for the first pair ($i=1$) through the last (n)

Because the formula is complicated, using graphing technology is helpful when dealing with large data sets. Calculating values manually is possible for smaller data sets.

- a. Use the data below — on unemployment rates overall and for people with some college education or an associate's degree in the four main regions identified by the Census Bureau — to complete the calculations table that follows. Note: Round your answers in the first two columns to two decimal places and those in the last three columns to four decimal places.

Unemployment Rates in Census Bureau Regions:
Overall vs. Some College Education or Associate's Degree

Region	Overall (%)	Some College Education or Associate's Degree (%)
Northeast	5.9	6.4
Midwest	5.5	5.5
South	6.0	5.9
West	6.4	6.5
	Mean = 5.950	Mean = 6.075

Calculations Table for Correlation Coefficients

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
5.9	6.4					
5.5	5.5					
6.0	5.9					
6.4	6.5					
Sums:				(A)		
				Square roots:	(B)	(C)

- b. With these values now identified, the formula for correlation coefficients can be simplified to:

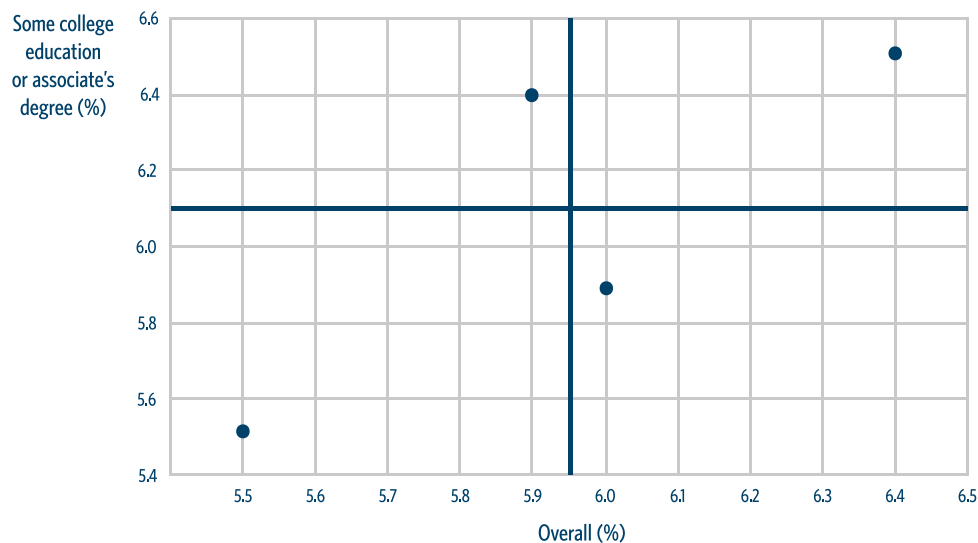
$$r = \frac{A}{B \cdot C}$$

A = the sum of all $(x_i - \bar{x})(y_i - \bar{y})$ values
 B = the square root of the sum of all $(x_i - \bar{x})^2$ values
 C = the square root of the sum of all $(y_i - \bar{y})^2$ values

Use the values for A, B, and C in the labeled boxes of your table to find the value of r , rounding to three decimal places.

- c. What does this r value indicate about the linear relationship between the two variables?
3. Calculate both $\sum_{i=1}^n (x_i - \bar{x})$ and $\sum_{i=1}^n (y_i - \bar{y})$. Why will these sums, representing the combined distance of each data point from the mean, always result in this same value?
4. The following scatter plot shows unemployment rates overall and for people with some college education or associate's degree for the four Census Bureau regions, with lines for the means of each variable (i.e., x and y) creating four quadrants. This setup helps you see the approximate distances of each data point from the means. (These distances are represented in the formula for r as $x_i - \bar{x}$ and $y_i - \bar{y}$.)

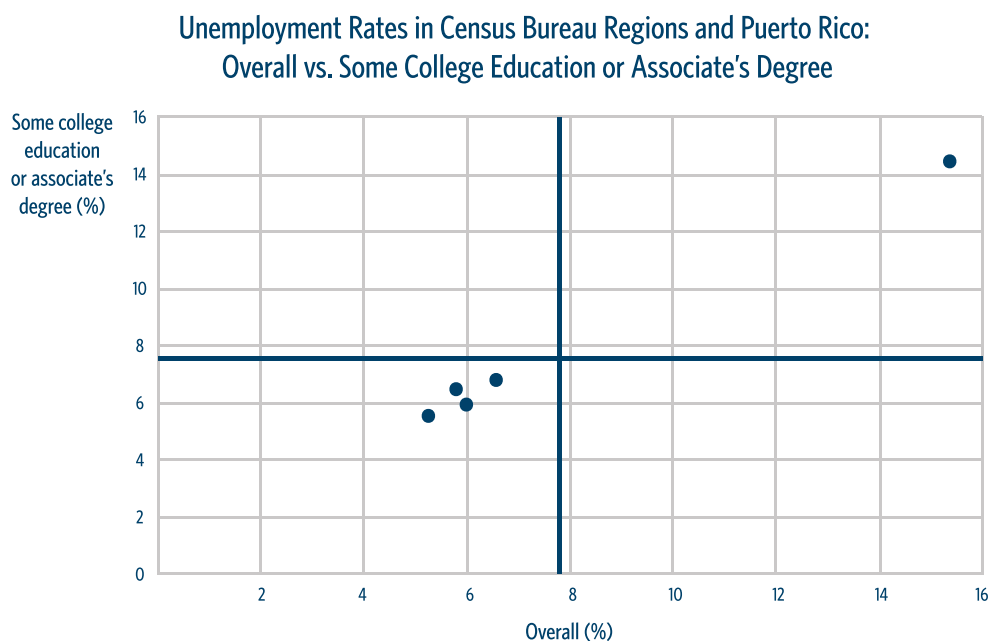
Unemployment Rates in Census Bureau Regions:
Overall vs. Some College Education or Associate's Degree



- a. In which quadrants do the points have a greater impact on the value of the correlation coefficient? Explain.

- b. In which quadrants will $(x_i - \bar{x})(y_i - \bar{y})$ be positive for this data set? Negative?
- c. How do you know that the sum of these values, represented in the formula's numerator as $\sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})]$, will be positive for this data set?
- d. Why will the denominator of the formula, or $\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}$, always be positive?
- e. Identify the largest and smallest possible values of r for any data set, and interpret what these values indicate for a linear relationship between two variables.
- Largest possible r value:
 - Smallest possible r value:

5. The following scatter plot includes the unemployment rates for Puerto Rico (15.2 percent overall; 14.3 percent for some college education or associate's degree), where the Census Bureau also collects data. The mean value lines have been adjusted accordingly.



- a. Do you think this new data point will have a large or small impact on the value of r ? Explain your logic.

- b. Test your prediction by recalculating the values in your calculations table and then determining the r value. Don't forget to first calculate the new x and y means.

$$\bar{x} = \quad \bar{y} =$$

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
5.90	6.40					
5.50	5.50					
6.00	5.90					
6.40	6.50					
15.2	14.3					
Sums:				(A)		
				Square roots:	(B)	(C)

$$r = \frac{A}{B \cdot C} =$$

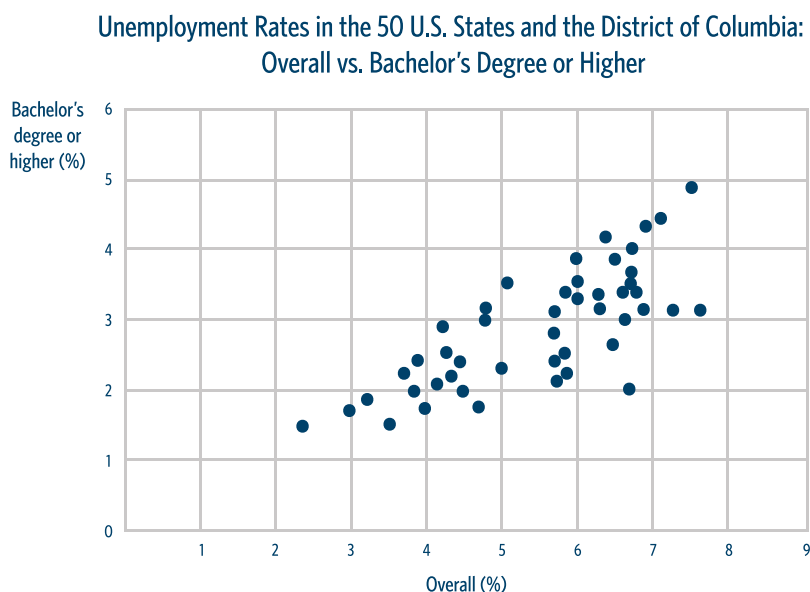
- c. How does your new r value compare with the original r value? What was the impact of adding the data for Puerto Rico to the original data set?

6. Some scatter plots contain influential points, which are data points that significantly affect both the slope of the line of best fit and the correlation coefficient. In regard to the correlation coefficient, a point is considered influential if it makes the association appear significantly stronger or weaker than it otherwise would be. Use mathematical evidence to explain why the data point for Puerto Rico would be considered an influential point.

Part 3 – Create Your Own Scatter Plots

1. Now use the data in **Item 1** to make your own scatter plot using graphing technology: Select at least 10 states (or the District of Columbia) and the unemployment rates for a specific education level, and plot them against the overall unemployment rates in each location you selected. Then use your scatter plot to answer the following questions and prompts.
 - a. Does there appear to be a relationship between unemployment rates for this educational attainment level and the overall rates for these locations? If so, describe the relationship as linear or not linear, as strong or weak, and as positive or negative.
 - b. Determine the correlation coefficient using graphing technology.
 - c. How does the value of the correlation coefficient, r , confirm or refute your interpretation in question 1a?

- d. If your scatter plot contains an influential point, remove it and then recalculate the correlation coefficient. How did that data point influence the correlation? Explain.
2. Next, use the data in **Item 1** for all 50 states and the District of Columbia to investigate the relationship between the unemployment rates of people with a bachelor's degree or higher and overall unemployment rates. Create a scatter plot with graphing technology and use it to answer the following questions:



- a. Does there appear to be a relationship between the unemployment rates for people with a bachelor's degree or higher and overall unemployment rates? If so, describe this relationship as linear or not linear, as strong or weak, and as positive or negative.
- b. Predict a value for the correlation coefficient.

- c. Using graphing technology, determine the value of the correlation coefficient. How close was your prediction?

- d. How does this actual value of r confirm or refute your interpretation in question 2a?

3. Based on what you have learned, would you agree or disagree that achieving higher levels of education makes you more employable? Explain, using specific data and the various analyses of your classmates from the discussion.

Part 4 – Draw Conclusions and Review Key Concepts

1. What conclusions can you draw from your and your classmates' investigations? Consider how the percentage of unemployed people at a given education level relates to the overall unemployment rate for a set of states, regions, or divisions.
2. Think back to the predictions you made before the activity. How accurate were they? Explain.
3. What is a correlation coefficient, and what can it tell you about the relationship between two variables?
4. Explain how to determine whether a data point on a scatter plot is influential.

Item 1: Unemployment Rates by Educational Attainment, 2014

Unemployment Rates by Educational Attainment Level for Each State and the District of Columbia (Aged 25 to 64)

	Overall (%)	Less Than High School Graduate (%)	High School Graduate or Equivalent (%)	Some College Education or Associate's Degree (%)	Bachelor's Degree or Higher (%)
Alabama	6.9	13.7	8.9	6.6	3.1
Alaska	6.7	14.0	12.0	5.8	2.0
Arizona	6.7	10.7	9.0	6.7	3.6
Arkansas	5.7	10.6	7.6	5.2	2.2
California	7.1	10.1	9.3	7.5	4.4
Colorado	4.3	7.5	5.4	4.7	2.9
Connecticut	6.6	13.9	9.0	7.1	3.8
Delaware	5.8	9.5	8.4	6.2	2.6
District of Columbia	7.7	14.6	17.6	14.6	3.1
Florida	6.9	12.1	8.7	6.5	4.3
Georgia	6.6	12.1	8.4	7.1	3.4
Hawaii	4.4	9.0	6.0	4.7	2.2
Idaho	4.5	8.9	5.6	4.4	2.4
Illinois	6.7	12.4	9.1	7.3	3.5
Indiana	5.7	11.9	7.4	5.5	2.5
Iowa	3.4	9.5	4.8	3.2	1.4
Kansas	4.2	8.3	5.9	4.3	2.1
Kentucky	6.3	13.7	7.9	5.8	3.1
Louisiana	6.0	12.4	6.6	5.8	3.2
Maine	5.0	11.1	7.0	5.2	2.3
Maryland	6.0	12.7	8.1	6.6	3.4

Item 1: Unemployment Rates by Educational Attainment, 2014 (Continued)

	Overall (%)	Less Than High School Graduate (%)	High School Graduate or Equivalent (%)	Some College Education or Associate's Degree (%)	Bachelor's Degree or Higher (%)
Massachusetts	5.8	11.9	8.6	6.6	3.3
Michigan	6.8	16.8	9.1	6.7	3.4
Minnesota	3.8	9.4	5.6	3.6	2.2
Mississippi	7.7	17.3	9.3	7.1	3.1
Missouri	5.7	14.3	7.5	5.6	2.8
Montana	4.0	10.0	5.8	4.1	1.6
Nebraska	3.2	8.4	4.5	2.8	1.8
Nevada	7.7	9.8	9.4	7.4	4.9
New Hampshire	4.3	8.3	6.2	4.5	2.5
New Jersey	6.4	10.4	8.6	7.0	4.1
New Mexico	7.3	13.6	9.1	7.4	3.1
New York	6.0	10.2	7.8	6.4	3.7
North Carolina	6.7	13.2	9.1	6.5	3.4
North Dakota	2.4	7.8	3.2	2.2	1.4
Ohio	5.8	15.5	7.0	5.9	2.6
Oklahoma	4.5	9.2	5.5	4.4	2.0
Oregon	6.7	13.7	8.5	6.6	4.0
Pennsylvania	5.7	13.1	7.0	6.0	3.1
Rhode Island	6.6	12.5	9.7	6.5	3.0
South Carolina	6.5	16.2	8.3	5.9	2.7
South Dakota	3.0	8.6	3.8	2.9	1.6
Tennessee	6.3	14.3	8.0	5.5	3.2

Item 1: Unemployment Rates by Educational Attainment, 2014 (Continued)

	Overall (%)	Less Than High School Graduate (%)	High School Graduate or Equivalent (%)	Some College Education or Associate's Degree (%)	Bachelor's Degree or Higher (%)
Texas	4.8	6.5	6.2	4.9	3.0
Utah	3.9	5.7	6.1	3.9	2.0
Vermont	4.7	12.0	7.2	4.9	1.9
Virginia	4.8	9.3	6.2	5.3	3.1
Washington	5.1	9.0	6.3	5.3	3.5
West Virginia	5.8	13.1	7.0	5.9	2.3
Wisconsin	4.5	10.5	5.5	4.6	2.4
Wyoming	3.9	12.8	4.5	3.3	2.4

[https://data.census.gov/cedsci/table?q=S2301%3A%20EMPLOYMENT 20STATUS&g=0100000US.04000.001&hidePreview=true&tid=ACST1Y2014.S2301](https://data.census.gov/cedsci/table?q=S2301%3A%20EMPLOYMENT%20STATUS&g=0100000US.04000.001&hidePreview=true&tid=ACST1Y2014.S2301)

To view the source data, copy and paste the URL above into your browser.

Item 1: Unemployment Rates by Educational Attainment, 2014 (Continued)

Unemployment Rates by Educational Attainment Level for Each U.S. Census Bureau Region (Aged 25 to 64)

	Overall (%)	Less Than High School Graduate (%)	High School Graduate or Equivalent (%)	Some College Education or Associate's Degree (%)	Bachelor's Degree or Higher (%)
Northeast	5.9	11.2	7.8	6.4	3.5
Midwest	5.5	12.7	7.2	5.5	2.7
South	6.0	10.6	7.7	5.9	3.3
West	6.4	10.0	8.3	6.5	3.9

https://data.census.gov/cedsci/table?q=S2301&hidePreview=true&tid=ACST1Y2014.S2301&vintage=2018&g=0100000US_0200000US2,1,3,4

To view the source data, copy and paste the URL above into your browser.

Item 1: Unemployment Rates by Educational Attainment, 2014 (Continued)

Unemployment Rates by Educational Attainment Level for Each U.S. Census Bureau Division (Aged 25 to 64)

	Overall (%)	Less Than High School Graduate (%)	High School Graduate or Equivalent (%)	Some College Education or Associate's Degree (%)	Bachelor's Degree or Higher (%)
New England	5.8	12.2	8.3	6.2	3.2
Middle Atlantic	6.0	10.9	7.7	6.4	3.6
East North Central	6.1	13.6	7.8	6.2	3.0
West North Central	4.2	10.6	5.8	4.0	2.1
South Atlantic	6.4	12.3	8.3	6.5	3.5
East South Central	6.7	14.5	8.4	6.1	3.1
West South Central	5.0	7.5	6.3	5.0	2.9
Mountain	5.6	9.6	7.4	5.6	3.1
Pacific	6.7	10.2	8.7	7.0	4.2

<https://data.census.gov/cedsci/table?q=S2301&hidePreview=true&tid=ACST1Y2014.S2301&vintage=2018&g=0300000US3,6,2,8,1,9,5,4,7>

To view the source data, copy and paste the URL above into your browser.